

Designing an Assessment Course for In-Service English Language Teachers in Colombia

Frank Giraldo¹  

Abstract

Research on language assessment literacy (LAL) has focused on describing this construct for different stakeholders involved in language assessment, while paying secondary attention to pedagogical initiatives aimed at fostering LAL among teachers. This research article reports on a diagnostic study which sought to describe the LAL of forty Colombian English language teachers; the purpose of this diagnostic was to use the teachers' feedback to design an online language assessment course. Through a mixed-methods research design, a questionnaire and a content analysis scheme were used for data collection and analysis. The convergent findings indicated issues in the design of assessment instruments and a related need to develop the teachers' design dimension of their LAL. Implications for the design of the course are presented, based on the data generated in the study.

Keywords: language assessment; language assessment literacy; language testing; teacher professional development; validity in language assessment

Desenvolvimento de um curso de avaliação para professores de inglês em serviço na Colômbia

Resumo

As pesquisas sobre avaliação em contexto de línguas (ACL) têm se concentrado em descrever esse constructo para diferentes atores envolvidos na avaliação de línguas, enquanto têm dado atenção secundária às iniciativas pedagógicas voltadas para o fortalecimento do ACL entre os professores. Este artigo de pesquisa relata um estudo diagnóstico que buscou descrever o ACL de quarenta professores colombianos de inglês; o objetivo do diagnóstico foi utilizar o feedback dos professores para projetar um curso online de avaliação linguística. Por meio de um desenho de pesquisa de métodos mistos, foram utilizados um questionário e um esquema de análise de conteúdo para a coleta e análise dos dados. Os achados convergentes indicaram problemas no desenho de instrumentos de avaliação e uma necessidade relacionada de desenvolver a dimensão de design do ACL dos professores. As implicações para o desenho do curso são apresentadas com base nos dados gerados no estudo.

Palavras-chave: avaliação de línguas; avaliação em contexto de línguas; testes de língua; desenvolvimento profissional docente; validade na avaliação de línguas

1 M. A. in Teaching English as a Second Language. Departamento de Lenguas Extranjeras – Universidad de Caldas en Manizales, Colombia. frank.giraldo@ucaldas.edu.co

Artículo de investigación

Para citar este artículo

Giraldo, F. (2025). Designing an Assessment Course for In-Service English Language Teachers in Colombia, *Folios*, (62), 33-49. <https://doi.org/10.17227/folios.62-20705>

Artículo recibido

24·01·2024

Artículo aprobado

28·01·2025

Artículo publicado

01·07·2025

Diseño de un curso de evaluación para docentes de inglés en servicio en Colombia

Resumen

La investigación sobre la Literacidad en Evaluación de Lenguas (LEL) se ha centrado en describir este constructo para diferentes actores involucrados en la evaluación lingüística, prestando una atención secundaria a las iniciativas pedagógicas destinadas a fomentar la LEL entre los docentes. Este artículo de investigación presenta un estudio diagnóstico que buscó describir la LEL de cuarenta docentes colombianos de inglés; el propósito de este diagnóstico fue utilizar la retroalimentación de los docentes para diseñar un curso en línea sobre evaluación de lenguas. A través de un diseño de investigación de métodos mixtos, se emplearon un cuestionario y un esquema de análisis de contenido para la recolección y el análisis de datos. Los hallazgos convergentes indicaron problemas en el diseño de instrumentos de evaluación y una necesidad relacionada de fortalecer la dimensión de diseño en la LEL de los docentes. Se presentan las implicaciones para el diseño del curso, basadas en los datos generados en el estudio.

Palabras clave: evaluación de lenguas; literacidad en evaluación de lenguas; pruebas de lengua; desarrollo profesional docente; validez en la evaluación de lenguas

Introduction

Language assessment literacy (henceforth LAL) refers to the knowledge, skills, and principles that various stakeholders (e.g., language testers, applied linguists, language teachers, etc.) need to participate in assessment-related tasks, namely the design and evaluation of language tests (Davies, 2008; Fulcher, 2012; Kremmel & Harding, 2020). Existing research on LAL has thoroughly described what this construct entails, even though it differs across stakeholder groups (Kremmel & Harding, 2020; Pill & Harding, 2013; Taylor, 2013). Regarding language teachers, scholars (Giraldo, 2022; Inbar-Lourie, 2013, November; Scarino, 2013; Stabler-Havener, 2018; Taylor, 2013) have suggested that teachers' LAL comprises, among other elements, the following:

- *Theoretical knowledge:* models and frameworks of language ability; language pedagogy; concepts in language testing.
- *Contextual knowledge:* personal beliefs, experiences, and needs; local assessment guidelines and practices.
- *Practical skills:* designing assessment instruments for all language skills; analyzing assessment data; evaluating language performance reliably and validly; connecting assessment to teaching and learning.
- *Principles:* ethics, fairness, democracy, and transparency.

Research on teachers' LAL has also focused on the skills and needs these stakeholders have (Fulcher, 2012; Sultana, 2019; Tsagari & Vogt, 2017; Vogt & Tsagari, 2014). Findings in this area have shown that, in general, teachers do not feel sufficiently prepared for assessment. As authors have suggested, this may be a result of inappropriate or non-existent training in language assessment at the pre-service level (Lam, 2015; Vogt & Tsagari, 2014). Thus, a research avenue that has been solidifying in later years is the development and implementation of LAL training for teachers.

Overall, as it may be expected, LAL training has had an overwhelmingly positive impact on teachers. Specifically, they develop an awareness of concepts related to language testing (Baker & Riches, 2018; Boyd & Donnarumma, 2018; Kremmel *et al.*, 2018); learn how to design professional assessment instruments (Koh *et al.*, 2018; Montee *et al.*, 2013; Restrepo, 2021); and analyze how assessment can positively impact teaching and learning (Arias *et al.*, 2012; Baker & Riches, 2018).

Although research has clearly shown the impact of LAL training on teachers, a current gap exists in research on how such LAL training initiatives—whether in the form of workshops or courses—have been designed. Montee *et al.* (2013), for example, mentioned that they used a questionnaire and the analysis of assessment instruments

to create an assessment course for teachers of less-commonly taught languages. Similarly, **Arias & Maturana (2005)** conducted a diagnostic study on language teachers' assessment practices and found problems regarding teachers' theoretical and practical knowledge for language assessment. The findings in this study led to a language assessment training initiative, reported in **Arias et al. (2012)**; however, **Arias et al.** did not discuss how such LAL training was designed.

Perhaps more importantly, research on the design of LAL courses could provide valuable feedback to advance LAL pedagogies—a current call in LAL research (**Fulcher, 2020; Giraldo, 2022**). Thus, the purpose of this article is to report on the design of an online language assessment course for in-service English language teachers in Colombia. Two related research questions guided the design of the course:

1. What are the language assessment literacy needs of a group of English teachers in Colombia?
2. How can these needs inform the design of an assessment course for these teachers?

This article begins with a literature review focusing on pedagogical approaches to fostering teachers' LAL. It then presents the research context and methodology, followed by findings and related discussions. The article concludes with final reflections and, most importantly, implications for the design of the online assessment course for the teachers in this study.

Literature Review

LAL Training for Teachers

Davies (2008) and **Malone (2017)** explain that a major source of LAL is language testing textbooks. As these authors suggest, the field has shown increasing concern with providing LAL for classroom teachers, which is evident in textbooks that directly target these stakeholders. Additionally, they explain that while there remains an emphasis on theoretical and technical aspects of testing, there is also growing attention to the critical and social dimensions of assessment—i.e., the principles side of LAL (e.g., ethics and fairness).

Language testing courses in undergraduate and graduate programs are another source of LAL reported in the literature. Findings in this aspect show that these courses, like textbooks, tend to focus on theoretical and practical aspects of assessment, while not placing sufficient emphasis on principles (**Brown & Bailey, 2008; Fulcher, 2012; Lam, 2015; O'Loughlin, 2006**).

The Nature of Assessment Courses for Language Teachers

A review of assessment courses for language teachers (for in-depth reviews, see **Giraldo, 2021; Gan & Lam, 2022**) suggests three trends in the literature. First, courses for teachers traditionally focus on assessment qualities such as *reliability* and *validity*, as well as the purposes that assessment serves (**Arias et al., 2012; Delgado & Rodriguez, 2022; Kleinsasser, 2005; Montee et al., 2013; Nier et al., 2009; O'Loughlin, 2006**). Second, most courses have a strong emphasis on analyzing assessment instruments to promote LAL, followed by the design of contextualized assessments (**Baker & Riches, 2018; Kleinsasser, 2005; Giraldo et al., 2023; Koh et al., 2018; Kremmel et al., 2018; Levi & Inbar-Lourie, 2020; Montee et al., 2013; Restrepo, 2021**). Finally, in conjunction with the practical aspect of these courses, teachers LAL is evaluated through the analysis of the test tasks and instruments they create (**Baker & Riches, 2018; Levi & Inbar-Lourie, 2020**).

As explained earlier, there are few studies reporting—in a detailed manner—on the actual design of assessment courses for language teachers. Thus, in the following sections, this article describes how an online language assessment course for in-service English teachers in Colombia was developed.

Methodology

A convergent mixed-methods approach was employed, whereby quantitative and qualitative data were collected to strengthen the findings and support informed decisions for course design. Mixed-methods approaches allow for thorough descriptions of phenomena, as data from various angles support broader and more nuanced interpretations (Edmonds & Kennedy, 2017; Ivankova & Greer, 2015). In the present study, qualitative data came from the analysis of assessment instruments that teachers shared, while the quantitative data were gathered through a questionnaire completed by the teachers. The research approach and findings from this study may provide useful insights for a larger audience: Other researchers interested in designing online assessment courses for language teachers may draw on these findings to inform their own LAL contexts.

Context of the Study

Various scholars in Colombia have called for the professional development of language teachers in LAL, at both the pre- and in-service levels (Giraldo & Murcia, 2019; Hernández-Ocampo, 2022; Herrera & Macías, 2015; López & Bernal, 2009). However, few studies report on LAL training specifically for in-service language teachers in this country. In fact, most existing research has focused on LAL courses for pre-service teachers (Giraldo & Murcia, 2019; Jaramillo-Delgado & Gil-Bedoya, 2019; Restrepo, 2020), although a growing body of research on LAL is emerging on in-service teacher training (Giraldo, 2024; Restrepo, 2021).

In response to the need to foster in-service teachers' LAL in Colombia, a project was proposed.² The project was divided into two major stages: 1) A diagnostic stage to collect data and design the LAL course and 2) an implementation stage to deliver it. The findings presented in this article are based on the diagnostic stage, which served as feedback for the design of the Online Language Assessment Course (henceforth OLAC).

To gather participants for the OLAC, information about it was emailed to all *Secretarías de Educación* (Offices of Education) in all 32 departments of Colombia. Additionally, the information was shared via academic contacts (professors and academics) and through personal social media (Facebook and WhatsApp). Any English language teacher in Colombia was eligible to participate, as OLAC was initially envisioned as a MOOC (Massive Open Online Courses).

A comprehensive, informative consent letter—available in both English and Spanish—was included in the call for participants to ensure that teachers were fully aware of the project's scope and the implications of their participation in the OLAC. Data collection took place in March 2023, and data analysis—used to inform the OLAC design—took place between May and July 2023.

Participants

Forty English language teachers from various regions of Colombia participated in the diagnostic stage of the aforementioned project. They were English language teachers in various educational contexts, ranging from elementary schools to universities. Figures 1 to 3 present information about participants' locations (by department), their years of experience teaching English, and the kinds of learners they taught at the time. Figure 4 provides information regarding teachers' prior training in language assessment.

2 The present article comes from a research study called *El Diseño y la Implementación de un Curso Online Masivo y Abierto (mooc) en Evaluación de Lenguas Extranjeras* (Designing and Implementing a Massive Open Online Course –mooc– on Language Assessment). The study was sponsored by the Vicerrectoría de Investigaciones y Posgrados of Universidad de Caldas in Manizales, Colombia. Code: 0727622.

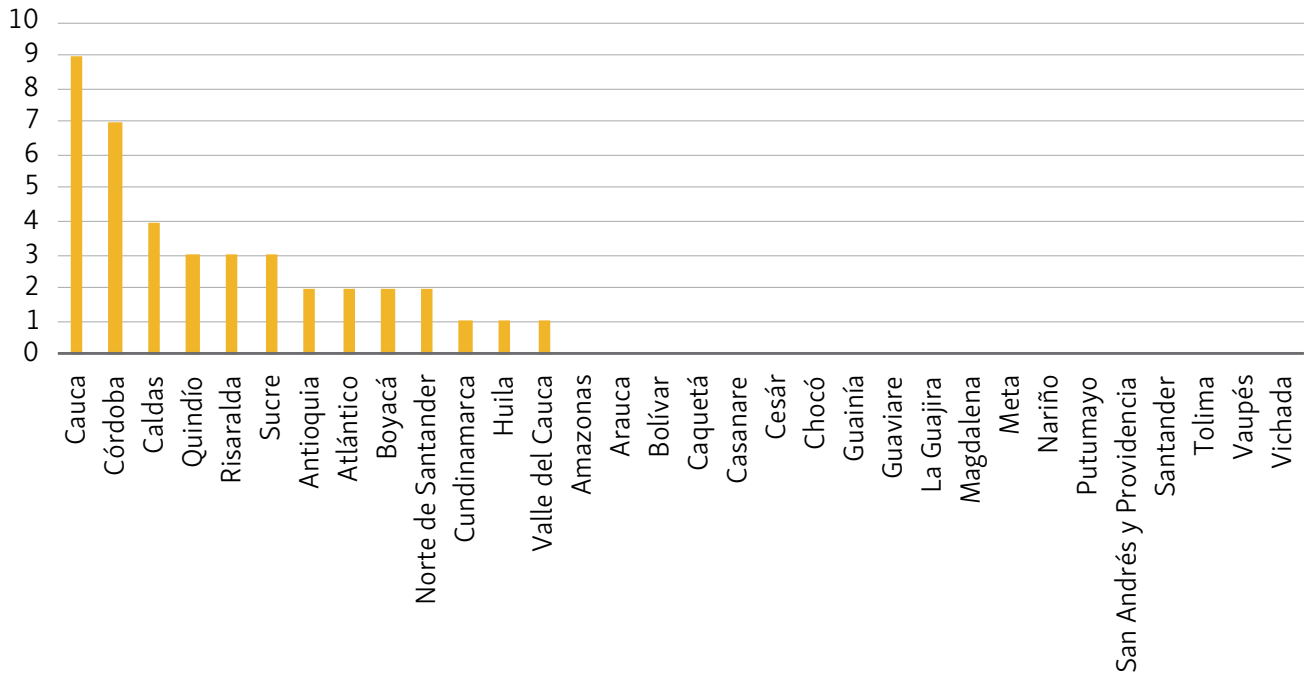


Figure 1. Number of Teachers Participating in Diagnostic Stage (per State)

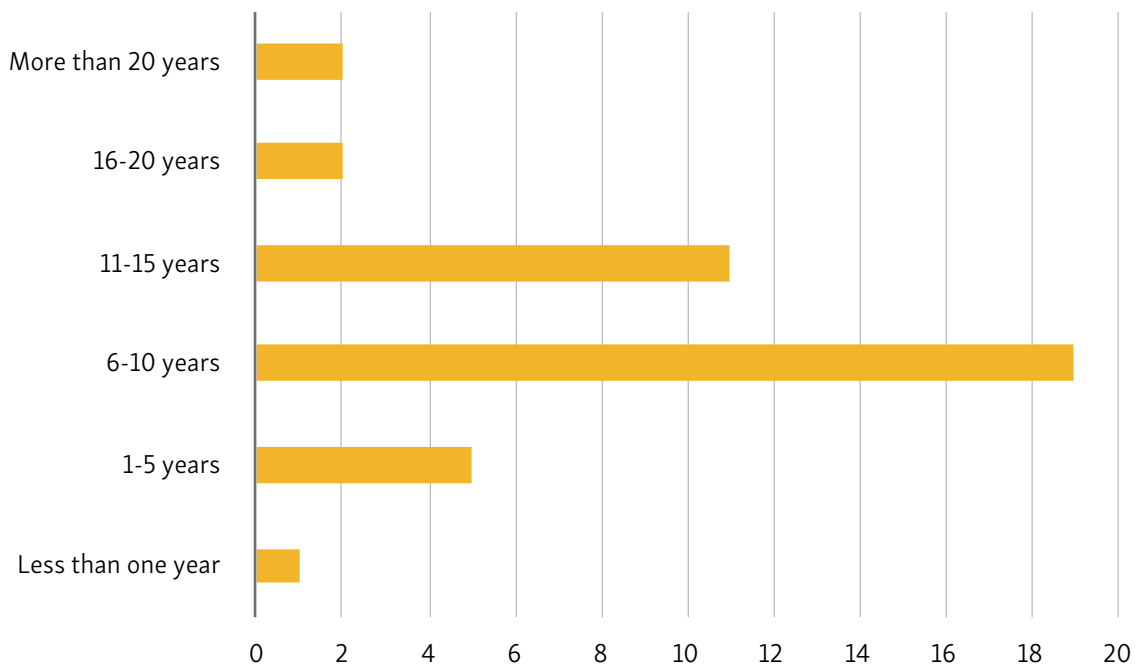


Figure 2. Experience Teaching English

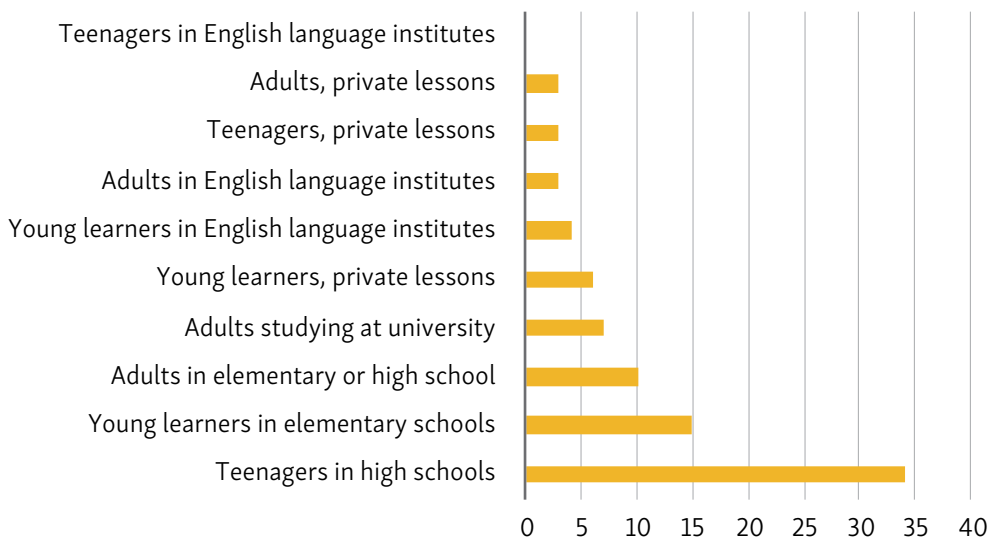


Figure 3. Types of Learners the Participants Teach (ranked)

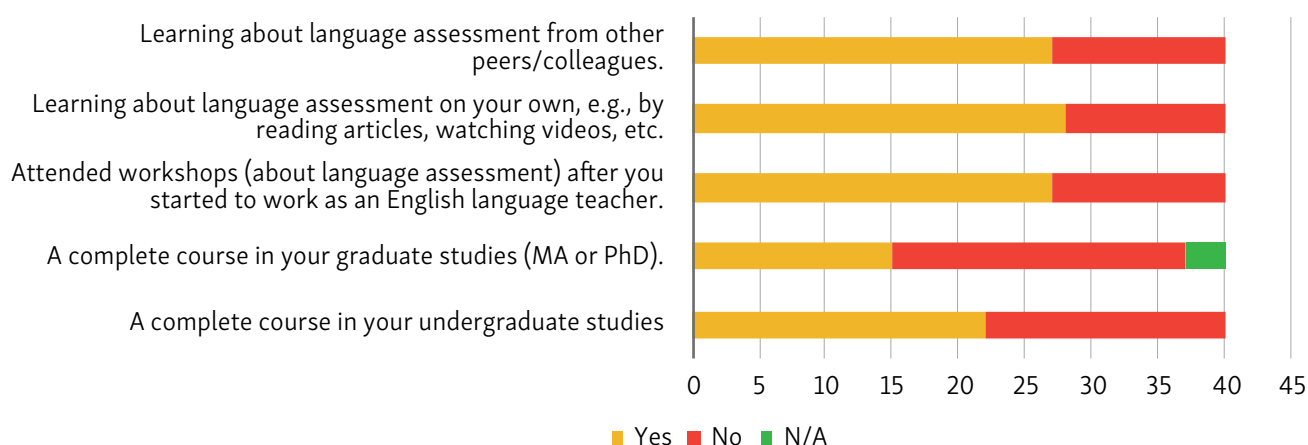


Figure 4. Prior Training in Language Assessment

Note: N/A means not applicable.

Corpus

To participate in the OLAC, the teachers were asked to share two assessment instruments: one for receptive skills (listening or reading) and one for productive skills (speaking or writing). In total, the teachers shared 80 documents, of which 40 were used in the final analysis.

Instrument and Procedures

Questionnaire. All participants completed an online questionnaire (designed using Google Forms), which was divided into five sections:

- Section 1 (9 items) gathered background information (e.g., name, email, place of work).
- Section 2 (8 items) focused on prior training in language assessment.

- Section 3 (9 items) asked teachers about the types of assessment activities in which they had been involved.
- Section 4 (3 open-ended items) explored teachers' learning goals, challenges in assessment, and expectations for the course.
- Section 5 (32 items) presented a list of suggested course topics—31 closed-ended and one open-ended item. The response scale ranged from 1 = Not important at all to 5 = Extremely important.

Appendix A includes the descriptive statistics for all the 31 closed-ended items in Section 5.

The questionnaire used in this study was based on [Giraldo & Yan \(2023\)](#) and piloted with 19 teachers. During the pilot, Cronbach's alpha was calculated at 0.92, which suggests a satisfactory level of internal consistency ([Dörnyei & Taguchi, 2010](#)). The teachers in the pilot exercise also provided comments on the usefulness of the questionnaire to design the OLAC. They agreed that the instrument was fit for purpose and that it included topics they found relevant for their assessment practice. One minor change was made to the questionnaire, based on the teachers' feedback: include an age-range that was missing in Section 1. In the final administration of this questionnaire with the 40 teachers, Cronbach's alpha was 0.90, which reiterated the satisfactory internal consistency of this instrument.

Analysis grid for the assessment instruments. A simple table was used to evaluate the strengths and areas for improvement in the assessment documents the teachers shared. This table included three columns: A code for each assessment instrument (e.g., Teacher1L), a section to record strengths, and a section to record aspects to improve. Analyzing language assessments at face value is not an easy task. To approach it objectively, a set of design guidelines was used, the most important of which are summarized below ([based on Alderson et al., 1995; Brown, 2011; Carr, 2011; Fulcher, 2010, among others](#)):

For receptive skills:

All the items in the reading/listening test

- have the potential to collect information about these skills and not others;
- do not overlap, i.e. one item does not give the answer to the other;
- are unambiguous and written in clear and correct language;
- do not have more than one answer in the multiple choice or true false formats, except where expressed otherwise in the instrument.

For productive skills:

The assessments for speaking/ writing

- include language-based criteria;
- do not have overlapping criteria;
- are based on a realistic or performance-based task;
- avoid the assessment of construct-irrelevant aspects;
- include clear scoring, where applicable.

Of the 80 documents shared, 40 instruments were appropriate for content analysis—that is, they were skills-based language assessments. [Table 1](#) shows how the 80 documents were categorized.

Table 1. Categorization of Assessment Documents

Documents Shared by Teachers Claiming to Test Receptive Skills (n = 40)	
Number of Documents	What the Tests Aimed to Assess
16	Listening
18	Reading
4	Grammar
1	Pronunciation (IPA transcription)
1	This was a lesson plan
Documents Shared by Teachers Claiming to Test Productive Skills (n = 40)	
Number of Documents	What the Tests Aimed to Assess
4	Speaking
2	Writing
9	Vocabulary
7	Grammar
3	Translation
2	These were lesson plans
9	Only task and no criteria
4	No document shared.

Data Analysis

The qualitative data in this study came from the content in the corpus of language assessments and the open-ended items in the questionnaire. To analyze these data, qualitative content analysis (QCA) was used. According to Schreier (2012), in QCA, a coding frame can be used, which is composed of main categories (or dimensions), subcategories, and units of coding. The frame can be arrived at through data-driven and concept-driven (theoretical) analysis. Table 2 is an example of how QCA was used for the data in the present research report.

Table 2. Approach for Qualitative Content Analysis

Main Category	Sub-Categories	Units of Coding
1. Design of language assessment instruments	1.1 Strengths in design	1.1.1 (from analysis of corpus) The test mixes methods, namely MCQs, T-F, and gap-filling.
	1.2 Aspects to improve in design.	1.1.2 (from analysis of corpus) The rubric includes clear descriptors.
		1.2.1 (from analysis of corpus) 4 out of 5 items can be answered without reading.
		1.2.2 (from questionnaire): What expectations do you have about this course? “Learn all the knowledge how to design good assessments”.

The quantitative data were analyzed using descriptive statistics, including mean, median, range, and ranks. These calculations were done for Section 5 of the questionnaire, which asked teachers to rate the importance of various topics to be included in the OLAC. In line with the mixed-methods approach for this case study, data from the questionnaire and the QCA were grouped to account for the major findings, which are presented and discussed in the next section.

Findings and Discussion

This section presents and discusses the findings emerging from qualitative and quantitative data on Colombian English teachers' feedback used to design an online language assessment course. To structure the report, results from the qualitative content analysis were presented first, followed by quantitative data from the questionnaire. Each finding will be followed by a discussion against relevant existing literature.

Issues with the Design of Receptive Skills Assessments

The analysis of the 34 assessment instruments for receptive skills led to the identification of design issues that could make the use of these instruments problematic for professional language assessment. Construct-irrelevance was a common design issue among the instruments, in various test methods. Various items and/or tasks for reading/listening seemed to be designed to assess other constructs unrelated to receptive or productive skills. For example, in Sample 1 below, the instrument is designed to assess how to use the words *more* and *than* in comparative sentences. Sample 2, on the other hand, was meant to assess listening comprehension; however, the item can be answered without listening as it assesses the ability to identify a synonym for a word in isolation.

Sample 1, from Teacher8R (gap filling):

Complete with *more* or *than*.

Stella thinks Simon's homework was _____ interesting _____ hers.

Sample 2, from Teacher20L (multiple choice):

A synonym for "Snooze" is

- Take a nap
- Scamper
- Sunny Back Porch

Another issue that was common in the instruments, particularly in multiple-choice and true-false items, was the fact that items did not have one unambiguously right answer, which they should. For example, in Sample 3, the teacher includes options that overlap.

Sample 3, from Teacher9L (multiple choice):

The 5 newly industrialised countries mentioned are:

- Brazil, China, Taiwan, South Korea and India
- India, China, Brazil, South Korea and Japan
- China, South Korea, Taiwan, Brazil and Vietnam
- India, China, Thailand, Brazil and South Korea

Let us suppose that the right answer is *d.* However, the countries mentioned in *d.* are also mentioned in *a.*, *b.*, and *c.* That means these other options are partially correct. Similarly, in Sample 4 below, the right answer is not clear as there is none. The item is asking for information that is not stated in the text students were supposed to read, so the item cannot be true nor false.

Sample 4, from Teacher11R (True-False)

Text to be read (only segment that mentions information about Gustavo):

My dad Gustavo isn't a doctor, he's a teacher. My mother is very good treating patients and my father is the best teaching English, Spanish and French.

Item:

Gustavo speaks German, Italian and Latin. True or False

According to the information in the text, and the item, one cannot state whether it is true or false that Gustavo speaks any of the three languages.

The teachers' answers in the questionnaire reiterated a need regarding the design of assessments for receptive skills. **Table 3** below shows the results from ranking 31 topics for the online assessment course, where assessing receptive skills was ranked third and creating test items was ranked eighth. Furthermore, notice that the range for assessing receptive skills is 1, which means the 40 teachers agreed on the importance of this topic for the course. The first ranked topic is included in the table as it directly relates to assessing receptive skills.

Table 3. Top Ranked Topics Related to Assessing Receptive Skills

Rank	Items	Means	Medians	Modes	Ranges
1	Planning the design of assessment instruments.	4.8	5	5	2
3	Assessing receptive skills: listening and reading.	4.75	5	5	1
8	Creating test items and tasks that can provide consistent and reliable information about your students' language skills.	4.575	5	5	3

Note: Table 3 only includes those rankings that are relevant for this finding.

Other research studies have suggested that teacher-made assessments of receptive language skills tend to have the same issues that are described above: construct irrelevance and ambiguity in answer keys (Arias & Maturana, 2005; Frodden et al, 2005; Giraldo et al., 2023). The design of useful reading or listening tests involves the adherence to numerous guidelines like the ones mentioned in the analysis grid above or discussed by authors (Alderson, 1995; Brown, 2011; Carr, 2011); thus, the teachers who designed these assessments may not have had the training needed to design useful tests of reading or listening comprehension (see Table 4 for details) or may have simply forgotten to follow specific design guidelines. Finally, based on the content analysis and the results in Table 3, the design of assessments for receptive language skills seemed to be a must for designing the online course.

Issues with the Design of Productive Skills Assessments

Whereas 34 assessment instruments of receptive skills were appropriate for analysis, six assessments of productive skills were analyzed: four speaking tests and two writing tests; three main reasons explain the low number of productive-skills tests. First, out of 40 assessment documents, 20 were tests of other language skills: nine tests sought to assess knowledge of vocabulary (e.g., providing the correct written translation for words in Spanish); seven tests targeted knowledge of grammar (e.g., forming syntactically correct sentences through unscrambling); three tests were exercises for students to translate sentences from Spanish into English; and two documents were lesson plans (see Table 1 for how the documents were classified). Second, nine assessments only included the task for learners to perform but did not specify assessment criteria in any way; thus, these nine documents, in reality, were tasks and not assessments. Finally, four teachers did not provide any document related to the assessment of productive language skills.

Issues with construct definition in speaking or writing were common across all six analyzed assessments. On the one hand, four tests included the criteria to be judged in either writing or speaking performances; however, there was no clear description of what was to be judged about the stated criteria. For instance, in Sample 5 below,

the teacher was seeking to assess writing skills within a tale-like genre. The assessment criteria were content, accuracy, vocabulary, and punctuation. However, the rubric does not specify what about these criteria was assessed; for instance, there were no details regarding what punctuation marks were the targets for this assessment task.

Sample 5, from Teacher15W

WRITING TEST

SCORE:

Read the *Canterville’s Ghost* again. Then, write a paragraph giving a different end to the story.

Grading criteria.:

Content: ___/10 | Accuracy: ___/10 | Vocabulary: ___/10 | Punctuation: ___/10

On the other hand, the other two tests included criteria that were unrelated to the construct of language ability; in other words, they assessed other aspects in addition to speaking or writing skills. Sample 6 is a rubric that was used to assess speaking; in this instrument, two criteria are not related to the construct of language ability or communicative competence: visual aids and preparedness. Let us suppose that, hypothetically speaking, a student is located at the *Superior* level in all the criteria. If this is the case, 50% of the performance will be attributed to these two constructs which are not language skills.

Sample 6, from Teacher25S

Criteria	Low	Basic	High	Superior
Visual Aids	Never uses visual aids during his/her presentation.	Uses simple visual aids such as random images that have no order, to support understanding of his/her presentation.	Uses organized visual aids that are relevant to the topic, to support understanding of his/her presentation.	Uses organized visual aids, whether physical or digital, with phrases that support understanding of his/her presentation.
Preparedness	Reads his/her entire presentation, showing little preparation for this activity.	Shows confidence while speaking and reading, in some cases, to present his/her idea.	Shows preparation when speaking during most of the presentation, using his/her notebook occasionally as an aid.	Is very prepared for the presentation and shows it by speaking without the need to use his/her notebook, classmates, or teacher.
Grammar				
Pronunciation				
Comments				

Note: The descriptors above were translated from Spanish. The descriptors for grammar and pronunciation were not included as they were not relevant to the present analysis and discussion.

In the open answers to the questionnaire, teachers reported that assessing productive skills and rubric creation were aspects they would expect to see in the course. For example, Teacher35 stated that they want to learn about “establishing criteria when assessing speaking”, whereas Teacher 21 reported that one of their challenges is “is how to evaluate the speaking skill.” Regarding rubric creation, Teacher17 stated that they “would like to learn more about qualitative evaluation, how to create rubrics”. Finally, assessing productive skills was ranked fourth, among the 31 topics in the questionnaire; the mean for this item was 4.7, the median 5, the mode 5, and the range 2, suggesting that the forty teachers found this topic very important for the course.

Other research studies have shown that, when it comes to assessing productive language skills, teachers do seem to have trouble clarifying the construct to be assessed (Arias *et al.*, 2012; Giraldo *et al.*, 2023; Levi & Inbar-Lourie, 2020).

As the data above suggests, this issue was also present in the instruments that were scrutinized. However, when it comes to including constructs unrelated to language (e.g., preparedness), two perspectives must be considered.

On the one hand, teachers in this study may need to include these other aspects, as part of institutional or even classroom policies for assessment. While these other criteria may pose validity issues for interpreting results from language assessments, Brookhart (2003) argues that in classroom assessment, context is construct-relevant. On the other hand, in general educational assessment, assessing skills unrelated to a class or course is called score pollution, which may be considered an unethical practice (Green et al., 2007; Rasooli et al., 2019); in language testing, this is called construct irrelevance and considered a threat to validity (Fulcher, 2010; Messick, 1989). In conclusion, the assessment of productive language skills must be addressed in the course, specific attention should be given to construct definition –and most importantly– the course should address to what extent assessing constructs unrelated to language ability is valid and ethical within teachers’ specific institutional and/or personal assessment ecosystems.

Instrument Design as a Major LAL Need

Especially in the open items of the questionnaire, the most common theme revolved around the need to learn or develop skills regarding the design of assessment instruments. This LAL need was expressed either as a learning want, challenge, or as a course expectation. The teachers used words such as tools or instruments to provide this feedback for designing the online course. For instance, the sample below shows how the teacher connects the design of assessments to language learning and development, within their educational context:

Teacher19

I would like to create tools that allow me to indisputably assess what my students learn as the foreign language lessons unfold. I also want to create tools to assess the students’ learning performance after they finish modules and lessons set in any foreign language learning course.

Additionally, Teacher22 comments on their challenge when it comes to test design. According to this teacher, a challenge for them is “establishing criteria when assessing speaking, writing and designing a listening test.” Notice that assessing productive skills comes up as a need again. Finally, Teacher31 states that they expect to “learn to create practical and reliable assessment tools”; this answer further suggests the teachers’ need to have a design-based assessment course.

The quantitative results from the questionnaire also may be interpreted as feedback to design a course in which the design of assessment instruments is a pillar. Table 4 presents three top ranked topics that involve the design of language assessments; the low range of 2 suggests that there was agreement among all the teachers, who thought these topics were very important for the course.

Table 4. Top Ranked Topics Related to Design as a Core Course Topic

Rank	Items	Means	Medians	Modes	Ranges
1	Planning the design of assessment instruments.	4.8	5	5	2
5	Creating test items and tasks that resemble real-life language use.	4.63	5	5	2
10	Designing test items which have the potential to collect clear information about your students’ language skills.	4.55	5	5	2

Note: Table 4 only includes those rankings that are relevant for this finding.

Perhaps not surprisingly, language teachers expect their learning and professional development to be grounded on a practical approach. This expectation has, in fact, been observed in LAL research studies which suggest that teachers want to learn about practical aspects in assessment (Fulcher, 2012; Malone, 2013). LAL courses, as the literature has shown, have responded to this call and have been based, mostly, on practical test analysis and test design tasks (see the section titled *The Nature of Assessment Courses for Language Teachers* for a review).

Limitations

Two limitations emerging from this research must be addressed. Initially, the project with which this report is aligned was called “*Designing a Massive Online Open Course (mooc) on Foreign Language Assessment*”. The goal was to create a course that would welcome a high number of Colombian English language teachers, which is why the invitation to participate was sent to the Offices of Education in all 32 states of Colombia. However, given that 40 teachers participated in the diagnostic stage, the MOOC became a small-scale assessment course. Therefore, the results of this sample are relevant to the ecology of this course and no generalizations can be made to other contexts. Notwithstanding the small nature of this research, the findings in this study may provide feedback for stakeholders interested in designing online assessment courses for language teachers elsewhere, more specifically, in the Colombian context of language teacher education.

The second limitation relates to the nature of the research design. The LAL of the 40 teachers in this study was studied through a questionnaire and content analysis of 80 assessment documents they shared, from which 40 instruments (34 for receptive skills and 6 for productive skills) were appropriate for scrutiny. The information from these two research methods can, in no way, provide a complete picture of teachers’ LAL, especially as this construct is developmental (Baker, 2021; Yan & Fan, 2021); thus, the data in this study, while practical and useful for LAL course design, is somewhat limited. Perhaps an in-depth, individual interview may have provided more data for course design and more nuances as to the teachers’ situated LAL. However, additional research methods may have made the initial project (the MOOC explained above) highly resource-heavy and time-consuming.

Conclusions and Implications

The purpose of the present research article was to report on the description of 40 Colombian English language teachers’ LAL. The information provided by these teachers became the basis for designing an online assessment course for these key stakeholders in LAL research. The convergent analysis of 40 assessment instruments and the results from the questionnaire were useful to reach two conclusions: on the one hand, there was a lack of clarity in the constructs to be assessed and how they were assessed; on the other hand, the assessment of constructs irrelevant to language ability was common. Therefore, as the data also shows, the professional design of assessments was a major need emerging from the data yielded by the two research methods.

Considering the data in this study, the online assessment course should be designed with these instructional principles in mind:

- Course sessions should address the design of assessment instruments for receptive and productive language skills.
- Guidelines for designing items to assess receptive skills are to be part of the content in the course; for example, how to make items construct relevant; how to design options that do not overlap; and how to spot faulty items and argue why they are faulty.
- Special consideration should be given to the theory and practice involved in designing useful rubrics to assess productive skills.

- The theoretical concept of validity should be addressed from a contextual, technical, and ethical perspective; in other words, teachers should be taught how to make their assessments more construct-relevant while considering their assessment contexts. Additionally, the course should touch upon how construct irrelevance may lead to unethical or unfair practices in assessment.

Appendix B is the schedule with the weekly topics for the OLAC. Some commentary accompanies the schedule to substantiate the decision-making process that bridged research and instruction for the present LAL initiative.

Appendix A

Descriptive Statistics for All Items in Section 5 of the Questionnaire

Rank	Items	Means	Medians	Modes	Ranges
1	Planning the design of assessment instruments.	4.8	5	5	2
2	Assessing integrated skills.	4.75	5	5	2
3	Assessing receptive skills: listening and reading.	4.75	5	5	1
4	Assessing productive skills: speaking and writing.	4.7	5	5	2
5	Creating test items and tasks that resemble real-life language use.	4.63	5	5	3
6	Relating language assessment to language teaching and learning.	4.6	5	5	3
7	Evaluating available resources (for example, technology) for test development.	4.58	5	5	2
8	Creating test items and tasks that can provide consistent and reliable information about your students' language skills.	4.58	5	5	2
9	Establishing a clear purpose for assessing your students' English.	4.58	5	5	4
10	Designing test items which have the potential to collect clear information about your students' language skills.	4.55	5	5	2
11	Using available resources efficiently during test development.	4.55	5	5	2
12	Evaluating the positive or negative influence that assessment can have on teaching and learning.	4.53	5	5	4
13	Assessing learners with special educational needs.	4.5	5	5	4
14	Conducting fair, ethical, and transparent assessment practices.	4.5	5	5	3
15	Administering assessments successfully.	4.5	5	5	2
16	Describing clearly the particular skills you want to assess.	4.5	5	5	2
17	Designing test tasks which have the potential to collect valid information about your students' language skills.	4.5	4.5	5	2
18	Evaluating whether a test has the potential to collect information about a determined set of language skills.	4.43	5	5	3
19	Assessing vocabulary.	4.43	5	5	2
20	Designing alternative assessments; for example, self- and peer assessment; portfolios.	4.43	5	5	2
21	Assessing pronunciation.	4.4	5	5	2
22	Evaluating whether an assessment instrument is meeting, or not, its purpose.	4.4	5	5	4
23	Assessing students in a bilingual mode.	4.38	4	5	3
24	Assessing different age groups (young learners, teenagers, etc.).	4.36	4.5	5	4
25	Analyzing misuses of language assessment.	4.33	4	4	3
26	Discussing ethics and fairness in classroom language assessment.	4.3	4	4	3
27	Assessing grammar.	4.3	4.5	5	2
28	Interpreting statistical information and scores from language assessment.	4.13	4	4	3
29	Evaluating large-scale or standardized language tests.	3.9	4	4	4
30	Designing traditional assessments, such as true-false and multiple-choice.	3.6	3.5	3	4
31	Learning about the history of language testing.	3.2	3	3	4

Appendix B

Schedule with Topics for the OLAC, with Commentary

Week	Content
1	Topic: Fundamentals of Language Assessment Commentary: Make emphasis on how purposes, methods, and constructs in language assessment must intertwine clearly.
2	Topic: Test Specifications Commentary: Remark on how test planning supports the design of useful, professionally made instruments.
3	Topic: Assessing Receptive Skills Commentary: Focus on guidelines for creating items that are construct-relevant and unambiguous.
4	Topic: Assessing Productive Skills Commentary: Place major emphasis on how carefully designed rubrics can be construct-relevant and lead to ethical practices, if used well.
5	Topic: Assessing Integrated Skills Commentary: Revisit design guidelines for receptive and productive skills assessments and place further emphasis on rubric creation.
6	Topic: Validity and Reliability Commentary: Draw attention to how miskeyed items or unclear/inexistent criteria may lead to reliability and validity problems; show how these two concepts are related in language assessment.
7	Topic: Review and Wrap-up Commentary: Base the final weeks of the course on feedback that the teachers give during the course, e.g., topics they would like to explore more in-depth.

References

- Alderson, C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.
- Arias, C. & Maturana, L. (2005). Evaluación en lenguas extranjeras: discursos y prácticas [Assessment in language teaching: Discourse and practices]. *Íkala, Revista de Lenguaje y Cultura*, 10(1) 63-91.
- Arias, C. I., Maturana, L. M. & Restrepo, M. I. (2012). Evaluación de los aprendizajes en lenguas extranjeras: hacia prácticas justas y democráticas [Assessment in foreign language learning: Towards fair and democratic practices]. *Lenguaje*, 40(1), 99-126. <https://doi.org/10.25100/lenguaje.v40i1.4945>
- Baker, B. (2021, June). *Unpacking the concept of language assessment literacy for all key stakeholders* [Conference presentation]. New Directions Colombia 2021 <https://americas.britishcouncil.org/new-directions/programme>
- Baker, B. A. & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557-581. <https://doi.org/10.1177/0265532217716732>
- Boyd, E. & Donnarumma, D. (2018). Assessment literacy for teachers: A pilot study investigating the challenges, benefits and impact of assessment literacy training. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 105-126). Springer. https://doi.org/10.1007/978-3-319-77177-9_7
- Brookhart, S. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12. <https://doi.org/10.1111/j.1745-3992.2003.tb00139.x>
- Brown, J. D. (2011). *Testing in language programs: A comprehensive guide to English language assessment*. McGraw Hill.
- Brown, J. D. & Bailey, K. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25(3), 349-383. <https://doi.org/10.1177/0265532208090157>
- Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford University Press.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347. <https://doi.org/10.1177/0265532208090156>
- Delgado, J. & Rodriguez, C. (2022). Language assessment literacy of language teachers in the context of adult education in Spain. *Studies in Language Testing*, 11(1), 64-91.
- Dörnyei, Z. & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing*. Routledge.

- Edmonds, W. A., & Kennedy, T. D. (2017). *An applied guide to research designs: Quantitative, qualitative, and mixed methods* (3rd ed.). SAGE.
- Fulcher, G. (2010). *Practical language testing*. Hodder Education.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Fulcher, G. (2020). Operationalizing assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8 -28). Cambridge Scholars.
- Gan, L. & Lam, R. (2022). A review on language assessment literacy: Trends, foci and contributions. *Language Assessment Quarterly*, 19(5), 503-525. <https://doi.org/10.1080/15434303.2022.2128802>
- Giraldo, F. (2021). Language assessment literacy and teachers' professional development: A review of the literature. *Profile: Issues in Teachers' Professional Development*, 23(2), 265–279. <https://doi.org/10.15446/profile.v23n2.90533>
- Giraldo, F. (2022). *Language assessment literacy and the professional development of pre-service foreign language teachers*. Editorial Universidad de Caldas.
- Giraldo, F. (2024). *Language assessment literacy for in-service teachers: Online course planning and implementation*. Editorial Universidad de Caldas.
- Giraldo, F., Naranjo-Trujillo, D. E. & Ariza-Villa, J. A. (2023). From the design of assessments to language assessment literacy. *Folios* (58), 128-141. <https://doi.org/10.17227/folios.58-16385>
- Giraldo, F. & Yan, X. (2023). Planning an online assessment course for English language teachers in Latin America. *Profile: Issues in Teachers' Professional Development*, 25(2), 147–164. <https://doi.org/10.15446/profile.v25n2.104703>
- Giraldo, F. & Murcia, D. (2019). Language assessment literacy and the professional development of pre-service language teachers. *Colombian Applied Linguistics Journal*, 21(2), 243–259. <https://doi.org/10.14483/22487085.14514>
- Green, S., Johnson, R., Kim, D. & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 2(7), 999–1011. <https://doi.org/10.1016/j.tate.2006.04.042>
- Hernández-Ocampo, S. P. (2022). Colombian scholars' discussion about language assessment: A review of five journals. *Profile: Issues in Teachers' Professional Development*, 24(2), 231–245. <https://doi.org/10.15446/profile.v24n2.94468>
- Herrera, L. & Macías, D. (2015). A call for language assessment literacy in the education and development of Teachers of English as a foreign language. *Colombian Applied Linguistics Journal*, 17(2), 302-312. <https://doi.org/10.14483/udistrital.jour.calj.2015.2.a09>
- Inbar-Lourie, O. (2013, November). *Language assessment literacy: What are the ingredients?* Paper presented at the 4th CBLA SIG Symposium Programme, University of Cyprus.
- Ivankova, N. & Greer, J. (2015). Mixed methods research and analysis. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (2nd edition pp. 63-81). Bloomsbury Publishing.
- Jaramillo-Delgado, J. & Gil-Bedoya, A. M. (2019). Pre-service English language teachers' use of reflective journals in an assessment and testing course. *Funlam Journal of Students' Research*, (4), 210–218. <https://doi.org/10.21501/25007858.3010>
- Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect*, 20(3), 77–102.
- Koh, K., Burke, L., Luke, A., Gong, W. & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, 22(3), 264–288. <https://doi.org/10.1177/1362168816684366>
- Kremmel, B., Eberharter, K., Holzknicht, F. & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high-stakes test development. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 173–194). Springer. https://doi.org/10.1007/978-3-319-77177-9_10
- Kremmel, B. & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120. <https://doi.org/10.1080/15434303.2019.1674855>
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. <https://doi.org/10.1177/0265532214554321>
- Levi, T. & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- López, A. & Bernal, R. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile: Issues in Teachers' Professional Development*, 11(2), 55-70

- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344. <https://doi.org/10.1177/0265532213480129>
- Malone, M. E. (2017). Training in language assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment: Encyclopedia of language and education* (3rd ed., pp. 225–240). Springer. https://doi.org/10.1007/978-3-319-02261-1_16
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Macmillan.
- Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, 13, 1–31.
- Nier, V., Donovan, A. & Malone, M. (2009). Increasing assessment literacy among LCTL instructors through blended learning. *Journal of the National Council of Less Commonly Taught Languages*, 9, 105-136.
- O'Loughlin, K. (2006). Learning about second language assessment: Insights from a postgraduate student on-line subject forum. *University of Sydney Papers in TESOL*, 1, 71–85.
- Pill, J. & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402. <https://doi.org/10.1177/0265532213480337>
- Rasooli, A., Zandi, H. & DeLuca, C. (2019). Conceptualising fairness in classroom assessment: exploring the value of organisational justice theory. *Assessment in Education: Principles, Policy & Practice*, 26(5), 584-611. <https://doi.org/10.1080/0969594X.2019.1593105>
- Restrepo, E. M. (2020). Monitoring preservice teachers' language assessment literacy development through journal writing. *Malaysian Journal of ELT Research*, 17(1), 38–52.
- Restrepo, E. (2021). Teacher educators' competence in language test design. In J. V. Abad (Ed.). *Research on language teaching and learning: Advances and projection* (pp. 165-191). Fondo Editorial Universidad Católica Luis Amigó.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327. <https://doi.org/10.1177/0265532213480128>
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE Publications.
- Stabler-Havener, M. L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics & TESOL*, 18(1), 1–22.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 2–14. <https://doi.org/10.1186/s40468-019-0077-8>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412. <https://doi.org/10.1177/0265532213480338>
- Tsagari, D. & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.
- Vogt, K. & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Yan, X. & Fan, J. (2021). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, 38(2), 219–246. <https://doi.org/10.1177/0265532220929924>